DOCUMENT RESUME

ED 346 148 TM 018 487

AUTHOR Marso, Ronald N.; Pigge, Fred L.
TITLE A Summary of Published Research: Classroom Teachers' Knowledge and Skills Related to the Development and Use of Teacher-Made Tests.
PUB DATE Apr 92
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classroom Techniques; *Elementary School Teachers; Elementary Secondary Education; *Knowledge Level; Literature Reviews; Meta Analysis; *Secondary School Teachers; Teacher Attitudes; Teacher Education; *Teacher Made Tests; *Test Construction; Testing Problems; Test Use
IDENTIFIERS *Teacher Competencies

ABSTRACT

This paper presents a summary of findings from a review of approximately 225 studies addressing the knowledge and skills of classroom teachers for kindergarten through grade 12 related to the development and use of teacher-made tests. The findings from the review suggest that little change in teachers' testing competence has occurred in the quarter century since S. T. Mayo first documented the following inadequacies in teachers' testing knowledge and training: (1) limited expertise, support, and preservice and inservice training are available to assist teachers in meeting their testing responsibilities; (2) teachers view teacher-devised testing as positively influencing instruction and learning; (3) most teacher-constructed tests contain many faults, and function almost exclusively at the recall level; and (4) teachers typically do not use test improvement strategies such as test blueprints or item analysis. Table 1 lists 32 practices, attitudes, and beliefs of teachers; and Table 2 summarizes 21 testing knowledge parameters and skills identified in the studies. There is a 109-item list of references. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED346148

A Summary of Published Research: Classroom Teachers' Knowledge and Skills Related to the Development and Use of Teacher-Made Tests

Ronald N. Marso and Fred L. Pigge
College of Education and Allied Professions
Bowling Green State University
Bowling Green, Ohio 43403

A paper presented at the Annual Meeting of the
American Educational Research Association
San Francisco, California

April 20-24, 1992

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
RONALD N. MARSO
FRED L. PIGGE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

This paper presents a summary of findings from a review of approximately 225 studies addressing K-12 classroom teachers' knowledge and skills related to the development and use of teacher-made tests. The findings from the review suggest that little change in teachers' testing competencies has occurred in the quarter century since Mayo first documented inadequacies in teachers' testing knowledge and training, that limited expertise, support, and preservice and inservice training are available to assist teachers in meeting their testing responsibilities, that teachers view teacher-devised testing as positively influencing instruction and learning, that most teacher-constructed tests contain many faults and function almost exclusively at the recall level, and that teachers typically do not use test improvement strategies such as test blue prints or item analysis.

TM018487

BEST COPY AVAILABLE

A Summary of Published Research: Classroom Teachers' Knowledge and Skills Related to the Development and Use of Teacher-Made Tests

Even though both teacher educators and measurement specialists commonly emphasize the significant contribution that teacher-made tests and related testing practices make to the classroom learning process (Crooks, 1988; Brophy & Good, 1986; Linn, 1990; Rosenshine, 1985), questions persist pertaining to the adequacy of teachers' training in and knowledge of testing and evaluation. For example, Gullickson (1986) has traced expressions of concerns about the adequacy of teachers' testing knowledge back to the professional literature of the early 1960's.

One probable reason for the inadequacy of teachers' testing knowledge may be that many teacher preparation programs do not require a testing and evaluation course of their teacher candidates. Several researchers (Gullickson & Hopkins, 1987; Roeder, 1973; Schafer and Lissitz, 1987) have gathered evidence which suggests that fewer than one-half of the educational institutions in our country require a testing and evaluation course for the preparation of teachers. Researchers have also reported that most educators avoid measurement courses when they are not required (Coffman, 1983; Stiggins & Bridgeford, 1982) and that teacher inservice training in testing and measurement is almost nonexistant (Dorr-Bremme, 1983; Gullickson, 1984).

The purpose of this paper is to provide a bibliography and selected findings from an extensive review of the research literature addressing K-12 classroom teachers' skills and knowledge related to the development and use of teacher-made tests. The full report of the findings from this review is scheduled to appear as a chapter in Teacher Training in Assessment, Steven Wise editor, in volume seven of the Buros Nebraska Symposium in Measurement and Testing. This paper presents information related to just three of the several questions addressed in the literature review: 1) What is the extent of classroom teachers' testing knowledge as revealed through their reported testing practices, beliefs, and attitudes? 2) What is the extent of classroom teachers' testing knowledge and skills as revealed through paper and pencil measures of teachers' testing knowledge and as revealed through ratings completed by teachers themselves or by their school supervisors or principals? 3) What is the extent of classroom teachers' testing knowledge and skills as revealed through direct analyses of samples of their teacher-made tests?

The research reports reviewed for the larger study were identified through computer searches of the ERIC data base and through reference citations within the computer identified studies. These procedures resulted in the collection of approximately 225 studies.

QUESTION ONE:

Teachers' Practices, Attitudes, and Beliefs

Much of what is known about teachers' tests and testing practices has been obtained through studies using teacher self-report data gathering procedures. These self-report studies provide a valuable but at best a limited understanding of teachers' actual testing knowledge and skills. Very few studies involving direct observations of teachers' testing practices or involving direct analyses

of teacher-constructed tests appear in the measurement literature. Consequently, little is known about what may be the true nature of classroom teachers' testing practices and the actual quality of their self-constructed tests (Stiggins, Conklin, & Bridgeford, 1986).

Teachers' Classroom Testing Practices

It has been estimated that a typical pupil will take between 400 and 1000 teacher-made tests before graduating from high school (Mehrens & Lehmann, 1987), that from 5 to 15 percent of a typical classroom day is devoted to some type of pupil assessment (Crooks, 1988; Haertel, 1986), and that teachers expend from 11 to 20 percent of a typical work day on some aspect of pupil assessment such as grading pupil work or the preparation, administration, and scoring of tests (Newman & Stallings, 1982; Stiggins, 1988). For example, in one study teachers reported constructing an average of 54.6 formal paper and pencil tests in a typical school year (Marso & Pigge, 1988a) as part of their many and diverse pupil assessment activities.

Teachers rely primarily on their self-constructed tests in assessing their pupils, but many teachers frequently use publisher-constructed (textbook or workbook) tests for this purpose as well. In one national sample of teachers, 95 percent reported using self-constructed tests and 77 percent reported using publisher-constructed tests (Dorr-Bremme, 1983). But regardless of the source of the test, it is clear that teachers and pupils spend considerable classroom time and effort in testing activities (Fleming & Chambers, 1983).

Teachers' testing practices have been found to vary somewhat by grade level of instruction and by subject area content being assessed. At the upper grade levels, teachers rely more on teacher-constructed as compared to publisher-constructed tests, express more concerns about the quality of pupil assessments, and are somewhat more likely to use test quality control procedures such as item analysis and checks on reliability than do teachers in the lower grades (Marso & Pigge, 1991; Stiggins & Bridgeford, 1985). Primary grade teachers place more focus upon assessment of pupil work samples than upon testing; lower elementary grade teachers more frequently use worksheets and tests provided in publisher textbooks and workbooks than do other teachers; and upper grade and high school teachers predominantly use formal self-constructed tests in their assessment of pupils (Herman & Dorr-Bremme, 1982; Salmon-Cox, 1981).

Essay questions are very seldom used by classroom teachers at any grade level. Although infrequently used, essay questions are more frequently found in English, history, and social studies tests than in other subject area tests; and they are more frequently used in the upper grades than in the lower grades.

Math and science teachers more frequently test their pupils as compared to other subject area teachers, and they rely more heavily upon paper and pencil tests than upon less formal assessment procedures. Teachers in writing and speech classes are more likely than are other teachers to depend upon direct observations and informal judgments than upon formal tests in assessing the progress of their pupils (Marso & Pigge, 1988a; Stiggins & Bridgeford, 1985).

Teachers in the upper grades tend to assign letter grades or marks based primarily on pupil test performance and daily work. In contrast, teachers in

the K-4 grades rely more on daily work and observations than upon tests in assigning grades. Nevertheless, teacher-made tests are considered to be at least one primary source of information about pupils for most teachers when assigning marks (Marso, 1986; Shulman, 1980).

Teachers rely more heavily on self-constructed tests as compared to other types of tests in their instructional practices, and they typically report constructing from 50 to 75 percent of the test questions used on their tests. Teachers also use a variety of test items with an average of 2.6 question types found on a typical teacher-devised test (Dorr-Bremme, 1983; Marso & Pigge, 1988a; Yeh, 1981).

Teachers most frequently use a combination of completion or short-response type questions in constructing their teacher-made tests followed by the use of matching, multiple-choice, true-false and essay type questions. When teachers were asked to rate the various item types on a single criterion described in terms of the usefulness, adaptability, and fairness to pupils, the question types are ranked from most useful to least useful in the following order: matching, completion, short-response, multiple-choice, true-false and essay. Although very infrequently used and perceived as not being very useful by most teachers, teachers believe both that pupils study more for essay tests as compared to objective tests and that essay tests are more likely to function at higher cognitive levels than are objective tests (Coffman, 1971; Marso, 1985).

Nearly all classroom teachers report that they provide pupils with feedback about their performance on tests following the administration of a classroom test, and typically they report spending about one-half of a class period for that purpose. Teachers also report that pupils usually are very attentive and motivated during these test feedback sessions (Haertel, 1986). Once teachers construct test questions, they tend to reuse them without analysis and revision, and, as noted previously, teachers report that they seldom use statistical procedures following the administration of a teacher-made test (Gullickson & Ellwein, 1985; Marso & Pigge, 1988c).

There are very few empirical studies revealing specifically how teachers use tests in their classroom instruction (Kuhs et al., 1985). Linn (1983), however, has described the linkage between classroom tests and instruction as involving four basic features: the match between test items and the instructional objectives, test provision of feedback for pupil performance and teacher instruction, the "flag" role of tests in pointing out key content to be studied, and the use of tests to assist in assigning pupil letter grades.

A number of survey investigations of teachers' testing practices have been conducted in the past decade. Generally, teachers report a heavy reliance on teacher-made tests in their day-to-day instruction; in contrast they report little reliance on standardized tests for making instructional decisions. Salmon-Cox (1981), after interviewing a sample of elementary teachers, reported that teachers made only minor use of the results from standardized tests in their classroom instruction, and Borg, Worthen and Valcarce (1986) reported unfavorable and indifferent classroom teacher attitudes toward the use of standardized tests but a highly positive attitude toward the use of teacher-made tests. Stiggins and Bridgeford (1985) reported that classroom teachers use their self-constructed tests for pupil diagnosis, grouping, grading, evaluation, and the reporting of pupil progress in their classrooms. These latter

researchers also reported that teachers placed more reliance on teacher-made tests than upon publisher-constructed tests (tests from workbooks, etc.), upon structured performance assessments, or upon spontaneous observations of pupils in making instructional decisions.

Dorr-Bremme (1983), following a survey of a national sample of school districts, revealed that the types of classroom assessments teachers rely on most heavily are characterized by immediate accessibility of scores, by an integration with teaching activities, and by a close tie between test questions and content taught. On each of these criteria standardized tests are at a disadvantage as compared to teacher-made tests. At all grade levels and for all criteria assessed, teachers in a study reported by Hall, Carroll and Comer (1988) attributed more value to teacher-prepared tests in making instructional decisions as opposed to standardized tests and as opposed to either district or state pupil minimum competency tests.

A persistent criticism of teachers is that they tend to over emphasize test scores, and in particular standardized test scores, relative to other available information about pupils. Hall, Carroll, and Comer (1988) found, however, that classroom teachers consistently favored the results of their self-constructed tests over the results of standardized or state competency tests in making decisions. Further, they noted that teachers made decisions with a reasonable regard for the complex data requirements of classroom settings. Similarly, Lazar-Morrison, Polin, Moy, and Burry (1980) concluded that teachers place greater confidence in the results of their own judgments of pupil performance than upon any formal tests; and Stiggins and Bridgeford (1985) reported that teachers rely on a number of sources of information in making decisions about pupils and that teachers' relative reliance on sources of pupil information is from highest to lowest the following: teacher-made tests, standardized tests, structured performance assessments, and spontaneous observations.

Other research related to the allegation that teachers over rely on test scores in making decisions about pupils also provides little support for this criticism of classroom teachers. Dorr-Bremme (1983) concluded that teachers bring several types of assessment information to their decisions about pupils but that they rely more on personal experiences and observations than upon test scores. Similarly, Salmon-Cox (1981) reported that high school teachers made very little use of standardized test scores in evaluating pupils; Shavelson, Cadwell and Izu (1977) found that teachers gave due consideration to the reliability of data in making decisions about pupils; and Kellaghan, Madaus, and Airasian (1982) found that teachers can accurately predict pupil test performance and only use students' standardized test scores to corroborate their own judgments.

More specifically, the findings of the research related to teachers' use of test scores in making decisions about pupils suggest that classroom teachers use scores to raise but not to lower their expectations of pupils. When teachers note a discrepancy between their perceptions of a pupil's ability and test scores, teachers tend to ignore test scores when the scores suggest that less might be expected of a pupil; whereas teachers tend to raise their expectations of a pupil when test scores suggest that more might be expected of a pupil (Airasian, Kellaghan, Madaus, & Pedulla, 1977).

Teachers' Attitudes and Beliefs about Testing

Although there is some inconsistency in the research findings about teachers' perceptions of their own testing ability, teachers typically: rate the effectiveness of their training in testing somewhat below the training they received in other professional areas (Gullickson, 1984; Marso & Pigge, 1987a), rate their testing proficiencies somewhat lower than their proficiencies in other professional knowledge or skill areas (Marso & Pigge, 1987a), and express concern about their testing skills and believe that they could benefit from practical training in tests and measurements skills (Crooks, 1988; Haertel, 1986). Relatedly, first-year teachers rank the extent of their concerns about pupil evaluation and assessment above all other professional concerns except for their concerns about classroom management, pupil motivation, and coping with individual differences among pupils (Veenman, 1984).

Teachers commonly do not feel confident about their ability to write good test questions (Carter, 1984; Gullickson, 1985; Stiggins & Bridgeford, 1985) and are uncertain about how to improve their tests (Carter, 1984). Teachers report that they believe many of their questions and concerns about testing could be alleviated through training (Carter, 1986). Conversely, researchers have reported that teachers express confidence in their tests as well as in their overall testing knowledge and do not want more training in testing (Green & Stager, 1986-87).

This apparent conflict in findings, which suggests that teachers seemingly both desire but do not want to partake in more training in testing, may have been explained at least in part by Stiggins (1988). He noted that teachers often do express confidence in their tests and in their general testing knowledge. Conversely, he stated that teachers are uncertain about technical aspects of testing and that teachers do want practical help in improving their tests and their testing practices. What teachers do not want, he concluded, is more of the theoretical-impractical training typically associated with tests and measurement courses and workshops.

Two studies of teachers' attitudes toward educational testing appear to be representative of teacher perceptions of tests and testing. Green and Stager (1986-87) surveyed 555 classroom teachers and reported that younger teachers are more skeptical of testing than older teachers, that upper grade teachers are more positive toward testing than are lower grade teachers who typically place more emphasis on classroom observations and informal pupil assessments rather than on formal tests, that teachers have a positive regard for teacher-made tests but tend to be negative or indifferent about standardized tests, that most teachers express interest in upgrading their testing skills, and that reported use of contemporary measurement practices (e.g., use of test specification tables and item analysis, etc.) was found to be somewhat related to more frequent pupil testing practices but not related to teachers' attitude toward testing.

In a second study of teachers' attitudes and beliefs about tests, Gullickson (1984) reported that teachers felt that teacher-constructed tests result in increased pupil effort, influence pupil self-concept, create desirable competition among students, improve interaction among pupils, improve the classroom learning environment, better focus teaching, provide a good learning experience for pupils, motivate pupil study, and accurately reveal pupil

progress. Further, Gullickson found that teachers believe that: frequent brief tests are more desirable than infrequent lengthy tests, school administrators encourage frequent testing of pupils, pupils prefer frequent tests, pupils try hard on tests, tests are an important instructional tool, tests need to be tied closely to instruction, tests help evaluate instruction, essay tests better assess pupil progress than objective tests and measure at a higher cognitive levels, tests should not be the sole determinant of pupil grades, and that tests are necessary to help justify grades to parents.

It may be that pupils reflect the attitudes of their teachers about tests, for students also feel that tests help them learn, and they too favor frequent testing. Pupils also report that teacher-made tests must be taken more seriously and are more difficult than standardized tests (Kulik & Kulik, 1981), and, like many teachers, some pupils feel that standardized tests are a waste of time (Stetz & Beck, 1981).

In summation, it appears that teachers expend considerable effort and time in fulfilling testing responsibilities in their classrooms; teachers schedule tests frequently followed by class discussions of pupil performance; teachers have concerns about but also positive feelings toward the role of testing and pupil evaluation in the instructional process; and teachers have confidence in their classroom tests and their overall testing ability but recognize that they would benefit from practical inservice training in testing. A more extensive listing of generalizations related to teachers' testing practices, attitudes, and beliefs is presented in Table 1.

QUESTION TWO:

Direct Assessments of Teachers' Testing Knowledge

As has been previously noted, very little research has been done involving the direct assessment of teachers' testing knowledge (Newman & Stallings, 1982). In this section of the paper brief descriptions of the findings from the very limited number of studies designed to directly assess teachers' testing knowledge, to rate the testing related proficiencies of teachers, and to directly assess teachers' test construction skills through analyses of their self-constructed tests are presented.

Among the earliest efforts to directly assess teachers' testing knowledge was the study reported by Mayo (1967). He conducted a large-scale national study sponsored by the National Council for Measurement in Education and funded by the U.S. Office of Education. In this study two forms of a test called Measurement Competency Test were administered to 2,877 graduating seniors in 86 teacher preparation institutions.

Mayo concluded from the teacher candidates' performance on the Measurement Competency Test that teacher training practices at that time had not sufficiently developed the levels of measurement competency of beginning teachers to assure their success in meeting testing and evaluation responsibilities demanded in classroom instruction. He recommended that preservice teacher measurement courses be improved, that a measurement course be compulsory for all teacher candidates, and that measurement courses have a practical focus in order to better reveal to preservice teachers their need of

measurement competencies and to better increase their commitment to attaining these competencies.

Mayo's testing of graduating college seniors (1967) and his survey of testing professionals (1964) have continued to be major reference points in the investigation of teachers' testing knowledge and skills, and the content of preservice measurement courses still reflect those topics deemed appropriate for the preparation of teachers by the testing professionals participating in his 1964 survey study. Providing further evidence of Mayo's continuing influence upon the measurement field, Newman and Stallings (1982) conducted what might be considered a follow-up of Mayo's study of teachers' testing knowledge. A battery of instruments patterned after Mayo's instruments, analyses of the content of several measurement textbooks, and a measurement item bank collected by the National Council for Measurement in Education were used by Newman and Stallings to assess the testing knowledge of teachers who were employed in three large southern school districts. A total of 294 K-12 inservice teachers identified through random selection procedures completed this battery of assessment instruments. Some of the findings from this study which relate to the purposes of this section of the paper follow (The percentages in parentheses are comparable figures from the Mayo study.):

1. Approximately 44 percent of the teachers in the sample had completed more training in measurement than one course, 33 percent (35%) had completed just one measurement course, about 6 percent (34%) took their measurement training as part of another course, and 13 percent (30%) had no formal measurement training.

2. The average percentage of questions answered correctly on the understanding of testing principles was 53.7 percent with teachers performing higher on general measurement principles than on technical aspects of testing.

3. As also was noted by Mayo, little difference in performance was found between teachers who had completed a testing course, with an average 54.6 percent correct response to the questions, and teachers who had not completed such a course, with an average 48.0 percent correct response.

4. The teachers in the sample reported making about one-half of their own tests and spent about 10 percent of their work time in testing activities.

5. The teachers in the sample reported greater use of objective than essay questions with most to least frequent use of question types as follows: completion, multiple-choice, matching, true-false, short answer, calculation, and essay.

6. It was concluded from the data collected that there had been little change in the unacceptable level of teachers' testing knowledge since Mayo's study in 1967, and these researchers, like Mayo, questioned the effectiveness of preservice teacher training in educational measurement.

Related, but less broadly based, studies tend to confirm the findings from the studies of Mayo and Newman and Stallings. Carter (1986) found that teachers were unaware of item writing faults or clues on a set of multiple-choice test questions even though a segment of their seventh grade pupils were sufficiently test wise to use the faults in answering the questions. Hills (1977) reported

that only 25 percent of the teachers in Florida showed adequate measurement preparation and that just 10 to 20 percent could correctly answer basic questions on educational measurement principles. Impara, Divine, Bruce, Liverman, and Gay (1990) found that classroom teachers had difficulty in answering questions related to scores derived from state mandated achievement tests. These researchers also reported that those teachers with formal measurement training scored somewhat higher than those teachers without formal measurement training (a mean difference of about one on a 17 item test) and that interpretive information designed to accompany the score reports increased teacher performance on the questions. Without the interpretive information 39 percent of the teachers answered fewer than 70 percent of the measurement questions correctly; whereas 10 percent of the teachers answered fewer than 70 percent of the measurement questions correctly with the information present.

In other studies Carter (1984) found that language arts teachers were unable to recognize the particular skill being measured by test questions, that teachers took more time and found it more difficult to construct test questions functioning at higher cognitive levels, and that teachers felt insecure about their knowledge of question writing principles and had previously spent little time editing and revising test questions. Finally, the findings from surveys of teachers' testing knowledge led Takeuchi (1977) and Infantino (1976) to conclude that teachers in California and New York, respectively, had rather superficial knowledge of tests and measurement.

In summation, the findings from studies utilizing direct assessments of teachers' tests and measurement knowledge suggest that teachers are not very knowledgeable about tests and measurement and that neither preservice nor inservice training appears to be rectifying the situation. Many practicing teachers report having received no formal measurement training during preservice training, many teachers report having received only a unit of measurement training as a part of another preservice course, and most teachers report having received no school sponsored inservice training or assistance in the development and use of tests (Dorr-Bremme, 1983).

Ratings of Teachers' Testing Proficiencies

Even though survey assessments of teachers' interests and skills are commonly used to help school administrators plan inservice instruction for teachers, just one study of this nature was located wherein the major focus was on the assessment of teachers' testing skills. Many other studies, however, collected and reported limited perceptual ratings of teachers' testing skills as secondary findings. The findings from these latter studies have already been reported in previous sections of this paper.

Marso and Pigge (1989a, 1989b, 1989c, 1988b, & 1987a) conducted a multifaceted statewide assessment of teachers' testing needs and proficiencies; findings from the various components of this study have been reported to audiences at different times and are referred to in different sections of this chapter. In this study teachers, principals, and supervisors rated classroom teachers' proficiencies in 26 testing skill areas. Approximately 320 classroom teachers with one to ten years of classroom teaching experience were asked to rate their current testing skill proficiencies; whereas the group of approximately 580 school principals and teacher supervisors were asked to rate the testing skill proficiencies of their typical beginning classroom teachers.

Additionally, teacher-constructed formal tests were collected from the teachers and were assessed for question types used, cognitive functioning levels, construction quality, etc.

Both teachers and administrators rated teachers' proficiencies in writing several types of test questions relatively low when compared to other proficiencies of teachers; whereas teachers' testing skills associated with pupil grading and test scoring, selecting good test questions, and appropriately handling the format of tests were rated relatively high by both groups. When teachers' actual tests were examined, however, it was found that question type writing skills rated highest in proficiency by the teachers and administrators were found to be the question types which violated more question writing guidelines, and the question writing skills rated lowest in proficiency by the teachers and administrators were found to violate fewer accepted question writing guidelines. In other words, a moderately high negative correlation was found between observed teachers' test question writing proficiencies and the teachers' and administrators' ratings of these same testing proficiencies (Marso & Pigge, 1989c).

The classroom teachers in this study also rated the effectiveness of their preservice teacher training in tests and measurement lower than the effectiveness of their total teacher training experience, lower than preparation received in their other education courses, and lower than the preparation received in their arts and science courses. Similarly, the administrators rated the testing and measurement proficiencies of their typical beginning teachers lower than they rated beginning teachers' knowledge of their subject areas, lower than they rated beginning teachers' other professional education proficiencies (e.g., instructional planning, handling discipline, etc.), and lower than they rated beginning teachers' overall proficiencies as educators.

QUESTION THREE:

Direct Assessments of Teacher-Made Tests

Rather surprisingly, very few studies of teachers' testing knowledge and skills have been conducted wherein direct analyses of samples of their teacher-made tests have served as the major data gathering procedure. One such study was reported by Fleming and Chambers (1983). They analyzed 342 teacher-made tests encompassing 8,800 test questions constructed by teachers assigned to several grade levels and subject areas in the Cleveland Public Schools. These tests and test questions were analyzed relative to Bloom's six cognitive functioning levels, question type use, subject content, grade level, and adherence to common question and format construction guidelines. Some of the more salient findings from this study follow:

1. Short-answer (including fill-in-the-blank) questions were most frequently used followed by matching, multiple-choice, true-false, and essay questions. True-false questions were infrequently used on the tests, and essay items were very infrequently used by these teachers' (about 1% of all questions).

2. Almost 80 percent of the questions found on the tests measured at the knowledge level. Approximately 94 percent of the questions on the junior high tests and 69 percent of the questions on all other tests examined were

judged to be functioning at the knowledge level. Rather than being spread equally throughout all the tests, the higher level functioning items, however, were found primarily on the math tests. Few questions on any tests were judged to measure pupils' ability to make applications.

3. Fewer than two-thirds of the tests contained directions for each question type.

4. Questions were grouped by question type on all tests, but questions often were not numbered consecutively and in some cases were not numbered at all.

5. Suggestive of inadequate support services, many of the tests were handwritten, were poorly reproduced, and had pages over-crowded with content. The combination of these factors were judged by the researchers to make many of the tests almost illegible.

6. Commonly identified question writing guideline violations included one or two word stems and illogical options in multiple-choice questions, matching items requiring fill-in-the-blank responses, and ambiguous short-answer response questions.

7. Most of the tests were approximately one or two pages in length and were comprised of approximately 30 questions with fewer questions present on the tests for the lower grades and more on the tests for the upper grades.

In a second broadly based study of a sample of teacher-made tests, Marso and Pigge (1991 & 1988a) analyzed 6504 test questions contained within 455 question exercises (a group of questions of similar type on a test) found on 175 formal teacher-made tests constructed by classroom teachers who had from one to 10 years of teaching experience; all of these teachers had completed a preservice tests and measurement course. These questions and tests were analyzed relative to item cognitive functioning levels using Bloom's six categories, violations of common test format and test question writing guidelines, question types and numbers of questions used, subject content measured, years of teachers' teaching experience, test grade level, and by setting of the school employing the teachers (urban, rural, and suburban). Some of the more salient findings from this study follow:

1. Question type use varied by grade level and subject area content. Essay questions were very infrequently (about 1% of all questions) used by all teachers and were least used by elementary level teachers. Elementary level teachers more frequently used completion and multiple-choice questions than did secondary teachers. Problem questions (calculation tasks) were the predominant question form used by math teachers; science teachers most commonly used multiple-choice, matching, and short-response questions; and English teachers most commonly used short-response and matching questions.

2. Very few differences were noted in test construction practices or test construction quality when the tests were classified by years of teachers' teaching experience and school setting.

3. Matching exercises were found to be the most error prone question type, and several question construction and test format construction guidelines were violated on many of the tests (e.g., eleven item writing type flaws appeared on 50% or more of the test exercises).

4. Teachers reported preparing an average of 54.6 formal teacher-made tests each year, approximately 70 percent of the teachers scheduled a test once every two weeks or more frequently in a typical class, and over 50 percent of the teachers reported writing three-fourths or more of the questions used on their tests.

5. The most frequently used question type used in the tests varied somewhat depending upon whether the criterion used was total number of questions or most frequently used question type exercise. The question types used from highest to lowest frequency were short-response, matching, true-false, multiple-choice, problems, completion, interpretive exercises, and essay. When total number of items, rather than how frequently this item type appeared on the tests sampled, were considered the items arranged from most frequent to least frequent in number were: multiple-choice, matching, short response, true/false, problems, completion, interpretive exerices, and essay.

6. As a total group of questions considering all tests, 72 percent were judged to be functioning at the knowledge cognitive level. When examined by test subject areas, however, this figure became more disturbing as a large majority of the questions functioning beyond the knowledge level were restricted to the math and science tests. In other subject areas the majority of the tests were found to be made up from 90 to 100 percent of the items judged to be functioning at the knowledge level.

7. Most teachers used a variety of test question types in their tests with an average of 2.6 question types per test.

In another study which involved the direct analysis of secondary teacher-constructed math and science tests, Oescher and Kirby (1990) analyzed 34 tests containing over 1400 test questions and gathered the responses of 35 teachers to a teacher testing practices questionnaire. These teachers reported that summative evaluation was the dominant purpose of classroom testing in actual practice, that they wrote over 65 percent of the questions used on their tests, that they were confident in their ability to construct good tests, that they used instructional objectives to guide their construction of test items, that they discussed pupils' test performance in class following an exam, and that they did not consistently use tables of test specification, item analysis procedures, or complete basic statistical analyses of their test scores such as the calculation of test score means and standard deviations. The direct analyses of these teachers' tests revealed that:

1. Format was in error on 70 percent of the tests (e.g., inadequate margins, spacing, etc.).

2. Directions were not present on 26 percent of the tests.

3. Over 60 percent of the questions were short-response questions with multiple-choice, matching, and true-false comprising 20, 15, and 5 percent of all questions, respectively.

4. Just four essay questions were present among the more than 1400 questions.

5. The teachers over estimated the number of their test items functioning beyond the knowledge level (Green, Halpin, & Halpin [1990] and Carter [1984] also noted this type of over estimation by teacher test writers). The teachers felt that about 25 percent of their questions measured beyond the knowledge and comprehension level, but judges determined the tests to contain an average of just eight percent of all questions measuring beyond the knowledge and comprehension levels. Somehat surprisingly, very few of the math test questions were judged to require pupils to apply knowledge of procedures to new situations.

6. All question types present on the tests were judged to violate several basic item writing guidelines (e.g., 17 of 18 multiple-choice exercises contained major flaws; whereas, short-response and true-false exercises were judged to be better constructed but still 50 percent of these question exercises contained construction flaws).

In other studies but where less comprehensive samples of teacher-made tests were examined, Billeh (1974) analyzed 33 science tests to determine their cognitive functioning levels and reported that of all questions reviewed 72 percent functioned at the knowledge level, 21 percent functioned at the comprehension level, and seven percent functioned at the application level. The more experienced teachers in Billeh's sample used more knowledge level items, but no differences in the cognitive functioning levels of the tests were found when classified by grade level or by extent of teacher training. Black (1980) reported an analysis of 48 secondary level science tests and found that the cognitive functioning levels of the tests varied between the science subject areas. Biology tests contained 94 percent knowledge, chemistry 66 percent knowledge, and physics 56 percent knowledge level questions.

Ball, Doss, and Dewalt (1986) studied the tests constructed by 74 junior and senior high social studies teachers. They found that, although approximately 75 percent of these teachers indicated that higher level instructional objectives were most important to student learning and approximately 25 percent of these teachers reported that they predominantly used these higher level type objectives in their teaching, 98 percent of the questions on these teacher-made social studies tests were measuring just at the recall level. Marso and Pigge (1988a) also found that the social studies tests collected in their study were composed of questions measuring almost exclusively 100 percent at the knowledge level.

Similarly, Stiggins, Griswold, and Wikeland (1989) conducted interviews, class observations, and direct analyses of teacher-constructed tests of 36 K-12 classroom teachers who had been participating in inservice teacher training focused on school district endorsed efforts to teach with a focus on the development of their pupils' thinking skills. They found that all of these teachers' self-constructed tests were composed of questions functioning 100 percent at the knowledge level except for the math tests. These researchers commented that it was easier to train teachers to teach with a focus on their

pupils' higher thinking levels than it was to train teachers to design tests to measure pupil achievement at these higher levels.

In summation, the review of studies of the ratings of teachers' testing proficiencies, of paper and pencil assessments of teachers' testing knowledge, and of direct analyses of teacher-constructed tests have provided further insight into teachers' testing knowledge, practices, and skills. School administrators and teachers, themselves, perceive teachers' proficiencies in testing skills to be somewhat below their other professional proficiencies. Paper and pencil testing of teachers' preservice and inservice knowledge about testing indicates that neither preservice nor inservice training in testing consistently results in individual teachers being knowledgeable about basic testing concepts and principles. And direct analyses of samples of teacher-made tests reveal frequent violations of the most commonly accepted question and test format writing guidelines and that teachers' self-constructed tests appear not to improve with increases in their years of teaching experience. A summary of more specific findings related to teachers' classroom testing knowledge derived from this review of studies of teachers' testing proficiencies, knowledge, and tests are presented in Table 2.

9/7

Bibliography

AERA-APA-NCME. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Airasian, P. W., Kellaghan, T., Madaus, G. F., & Pedulla, J. J. (1977). Proportion and direction of teacher rating changes of pupil progress attributable to standardized test information. Journal of Educational Psychology, 69(6), 702-709.

Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. Journal of Educational Measurement, 20(2), 103-119.

Balch, J. (1964). The influence of the evaluating instrument on students' learning. American Educational Research Journal, 1, 169-182.

Ball, D. W., Doss, A. R., & Dewalt, M. W. (1986). Level of teacher objectives and their classroom tests: Match or mismatch. Journal of Social Studies Research, 10(2), 27-31.

Bangert-Drowns, R., Kulik, J., & Kulik, C. (1988). Effects of Frequent Classroom Testing. Unpublished manuscript, Univeristy of Michigan.

Billeh, V. Y. (1974). An analysis of teacher-made test items in light of the taxonomic objectives of education. Science Education, 58, 313-319.

Black, T. R. (1980). An analysis of levels of thinking in Nigerian science teachers' examinations. Journal of Research in Science Teaching, 17, 301-306.

Borg, W. R., Worthen, B. R., & Valcarce, R. W. (1986). Teachers' perceptions of the importance of educational measurement. Journal of Experimental Education, 55(1), 9-14.

Bridgeman, B. (1974). Effects of test score feedback on immediately subsequent test performance. Journal of Educational Psychology, 66, 62-66.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Whittrock (Ed.) Handbook of Research on Teaching (3rd ed., pp. 328-374). New York: Macmillan.

Bushway, A., & Nash, W. (1977). School cheating behavior. Review of Educational Research, 147, 623-632.

Carrier, C., & Titus, A. (1981). Effects of notetaking pretraining and text mode expectations on learning from lectures. American Educational Research Journal, 18, 385-397.

Carter, K. (1984). Do teachers understand the principles for writing tests? Journal of Teacher Education, 35, 57-60.

Carter, K. (1986). Test-wiseness for teachers and students. Educational Measurement: Issues and Practice, 5, 20-23.

BEST COPY AVAILABLE

Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed., pp. 271-302). Washington, DC: American Council on Education.

Coffman, W. E. (1983). Testing in the schools: A historical perspective. Paper presented at the annual invitational conference of the Center for the Study of Evaluation, University of California, Los Angeles.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58, 438-481.

Diamond, E. E., & Fremer, J. (1989). The joint committee on testing practices and the code of fair testing practices in education. Educational Measurement: Issues and Practices, 8, 23-27.

Dorr-Bremme, D. W. (1983). Assessing students: Teachers' routine practices and reasoning. Evaluation Comment, 6, 1-12.

D'Ydewalle, G., Swerts, A., & DeCorte, E. (1983). Study time and test performance as a function of test expectations. Contemporary Educational Psychology, 8, 55-67.

Elton, L., & Laurillard, D. (1979). Trends in research on student learning. Studies in Higher Education, 4, 87-102.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E Hathaway (Ed.), Testing in the Schools: New Directions for Testing and Measurement, (No. 19, pp. 29-38). San Francisco: Jossey-Bass.

Frisbie, D. A., & Friedman, S. J. (1987). Test standards: Some implications for the measurement curriculum. Educational Measurement: Issues and Practice, 6(3), 17-23.

Gay, L. R. (1980). The comparative effects of multiple-choice versus short-answer tests on retention. Journal of Educational Measurement, 17, 45-50.

Green, K. E., & Stager, S. F. (1986-87). Testing: Coursework, attitudes, and practices. Educational Research Quarterly, 11(2), 48-55.

Green, K. E., & Williams, J. E. (1989). Standardized test use by classroom teachers: Effects of training and grade level taught. Paper presented at the National Council on Measurement in Education annual conference, San Francisco.

Green, S. B., Halpin, G., & Halpin, G. (1990). Classroom tests assess higher-order cognitive functioning: An illusion of college instructors. Paper presented at the National Council on Measurement in Education, Boston.

Gronlund, N. E., & Linn, R. L. (1990). Measurement and Evaluation in Teaching (6th ed.). New York: Macmillan.

Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77, 244-248.

Gullickson, A. R. (1985). Student evaluation techniques and their relationship to grade and curriculum. Journal of Educational Research, 79(2), 96-100.

Gullickson, A. R. (1986a). Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23, 347-354.

Gullickson, A. R. (1986b). Perspective on educational measurement. Applied Psychology Measurement, 10, 109-132.

Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, 4(1), 15-18.

Gullickson, A. R., & Hopkins, K. D. (1987). The context of educational measurement instruction for preservice teachers: Professor perspectives. Educational Measurement: Issues and Practice, 6, 12-16.

Guza, D. J., & McLaughlin, T. F. (1987). A comparison of daily and weekly testing on student spelling performance. Journal of Educational Research, 80(6), 373-376.

Haertel, E. (1986). Choosing and using classroom tests: Teachers' perspectives on assessment. Paper presented at the American Educational Research Association, San Francisco.

Hall, B. W., Carroll, D., & Comer, C. B. (1988). Test use among classroom teachers and its relationship to teaching level and teaching practices. Applied Measurement in Education, 1(2), 145-156.

Halpin, G., & Halpin, G. (1982). Experimental investigation of the effects of study and testing on student learning, retention, and ratings of instruction. Journal of Educational Psychology, 74(1), 32-38.

Herman, J., & Dorr-Bremme, D. (1982). Assessing students: Teachers' routine practices and reasoning. Paper presented at the annual meeting of the American Educational Research Association, New York.

Hermanowicz, H. J. (1980). Toward a new model of teacher education. Paper presented at the Crossgate Seminar sponsored by the Pennsylvania Department of Education, Mechanicsburg, PA. (ERIC Document Reproduction Service No. ED 199215)

Hill, K. T., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. Elementary School Journal, 85, 105-126.

Hills, J. R. (1977). Coordinators of accountability view teachers' measurement competence. Florida Journal of Education Research, 19, 34-44.

Impara, J. C., Divine, K., Bruce, F. A., Liverman, M., & Gay, A. (1990). Utility of interpretive information in helping teachers interpret standardized test score reports: The Virginia state assessment program. Paper presented at the National Council on Measurement in Education, Boston.

Infantino, R. L. (1976). Testing and accountability: A survey of the knowledge and attitudes of New York State secondary school English teachers. Doctoral dissertation, State University of New York at Buffalo. (University Microfilms No. 77-6143)

Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). The effects of standardized testing. Boston: Kluwer-Nijoff.

Kinney, D., Brickell, J., & Lynn, M. A. (1988). Testing programs in Illinois elementary school districts: The relationships between standardized achievement, competency-based, and teacher-made tests. Paper presented at the American Educational Research Association annual meeting, New Orleans.

Klimko, I. P. (1984). Item arrangement, cognitive entry characteristics, sex, and test anxiety as predictors of achievement examination performance. Journal of Experimental Education, 52(4), 214-219.

Kuhs, T., Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1985). Differences among teachers in their use of curriculum embedded tests. The Elementary School Journal, 86, 141-153.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. Review of Educational Research, 58(1), 79-97.

Lambert, R. F. (1980-81). Teacher attitudes on testing: A multiple perspective. College Board Review, 118, 13-14, 29-30.

Lazar-Morrison, C., Polin, L., Moy, R., & Burry, J. (1980). A review of the literature on test use. CSE Report No. 144. Los Angeles, CA: Center for the Study of Evaluation.

Linn, R. L. (1983). Testing and instruction: Links and distinctions. Journal of Educational Measurement, 20, 179-189.

Linn, R. L. (1990). Essentials of student assessment: From accountability to instructional aid. Teachers College Record, 91(3), 422-436.

Marso, R. N. (1969). The influence of test difficulty upon study efforts and achievement. American Educational Research Journal, 6(4), 621-632.

Marso, R. N. (1970a). Classroom testing procedures, test anxiety, and achievement. The Journal of Experimental Education, 38(3), 54-58.

Marso, R. N. (1970b). Test item arrangement, testing time, and performance. Journal of Educational Measurement, 7(2), 113-118.

Marso, R. N. (1985). Reliability and validity of essay and objective test items in a classroom setting. Educational Review, 9(1), 9-20.

Marso, R. N. (1986). Testing practices and test item preferences of classroom teachers. Paper presented at the Midwestern Educational Research Association, Chicago. (ERIC Document: ED 268 145)

Marso, R. N., & Pigge, F. L. (1987a). A state-wide assessment of the testing and evaluation needs and proficiencies of beginning teachers: Implication for staff development. A paper presented at the Association of Supervision and Curriculum Development, New Orleans. (ERIC Document: ED 283 833)

Marso, R. N., & Pigge, F. L. (1987b). Differences between self-perceived job expectations and job realities of beginning teachers. Journal of Teacher Education, 38(6), 53-56.

Marso, R. N., & Pigge, F. L. (1988a). An analysis of teacher-made tests: Testing practices, cognitive demands and item construction errors. A paper presented at the National Council on Measurement in Education, New Orleans. (ERIC Document: ED 298174)

Marso, R. N., & Pigge, F. L. (1988b). Ohio secondary teachers' testing needs and proficiencies: Assessments by teachers, supervisors, and principals. American Secondary Education, 17(2), 2-9.

Marso, R. N., & Pigge, F. L. (1988c). Teacher-made tests and testing: Classroom resources, guidelines and practices. A paper presented at the Midwestern Educational Research Association, Chicago. (ERIC Document: ED 291 781)

Marso, R. N., & Pigge, F. L. (1988d). The availability in Ohio's schools of resources to support teacher-made testing: Surely we can do better! OASCD Dimensions, 9(5), 8-9.

Marso, R. N., & Pigge, F. L. (1989a). Elementary classroom teachers' testing needs and proficiencies: Multiple assessments and inservice training priorities. Educational Review, 13, 1-17.

Marso, R. N., & Pigge, F. L. (1989b). Standardized and competency testing programs: Classroom teachers' needs and proficiencies as viewed by principals and supervisors. Paper presented at the Midwestern Educational Research Association, Chicago. (ERIC Document: ED 304 473)

Marso, R. N., & Pigge, F. L. (1989c). The status of classroom teachers' test construction proficiencies: Assessment by teachers, principals, and supervisors validated by analyses of actual teacher-made tests. A paper presented at the National Council on Measurement in Education, San Francisco. (ERIC Document: ED 306 283)

Marso, R. N., & Pigge, F. L. (1990). Training, job titles, and responsibilities of directors of public school standardized testing programs. Paper presented at the National Council for Measurement in Education, Boston, MA.

Marso, R. N., & Pigge, F. L. (1991). An analysis of teacher-made tests: Testing practices, cognitive demands and item construction errors. Contemporary Educational Psychology, 16, 279-286.

Mayo, S. T. (1964). What experts think teachers ought to know about educational Measurement. Journal of Educational Measurement, 1(1), 79-86.

Mayo, S. T. (1967). Preservice preparation of teachers in educational measurement. (Contract No. OE 4-10-011). Chicago, IL: Loyola University.

Mehrens, W. A., & Lehmann, I. J. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York: Holt, Rinehart and Winston.

Mehrens, W. A., & Lehmann, I. J. (1987). Using standardized tests in education (4th ed.). New York: Longman.

Monk, J. J., & Stallings, W. M. (1970). Effects of item order on test scores. Journal of Educational Research, 63, 463-465.

Monk, J. J., & Stallings, W. M. (1971). Another look at the relationship between frequency of testing and learning. Science Education, 55(2),. 183-188.

Newman, D. C., & Stallings, W. M. (1982). Teacher competency in classroom testing, measurement preparation, and classroom testing practices. Paper presented at the American Educational Research Association annual meeting, New York. (ERIC Document: ED 220 491)

Newman, D. L., Kundert, D. K., Lane, D. S., & Bull, K. S. (1988). Effect of varying item order on multiple-choice test scores: Importance of statistical and cognitive difficulty. Applied Measurement in Education, 1(1), 89-97.

Nickerson, R. S. (1989). New directions in educational assessment. Educational Researcher, 18, 3-7.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. Journal of Educational Psychology, 74, 18-22.

Oescher, J., & Kirby, P. C. (1990). Assessing teacher-made tests in secondary math and science classrooms. Paper presented at the National Council on Measurement in Education, Boston.

Peckham, P. D., & Roe, M. D. (1977). The effects of frequent testing. Journal of Research and Development in Education, 10, 40-50.

Roeder, H. H. (1973). Teacher education curricula--your final grade is F. Journal of Educational Measurement, 10(2), 141-143.

Rogers, E. M. (1969). Examinations: Powerful agents for good or ill in teaching. American Journal of Physics, 37, 954-962.

Rosenshine, B. (1985). Teaching functions in instructional programs. The Elementary School Journal, 83(4), 335-351.

Ruddell, R. B. (1985). Knowledge and attitudes toward testing: Field educators and legislators. The Reading Teacher, 38, 538-543.

Rudman, H. C., Kelly, J. L., Wanous, D. S., Mehrens, W. A., Clark, C. M., & Porter, A. C. (1980). Integrating assessment with instruction: A review (1922-1980) (Research Series No. 75). East Lansing, MI: Michigan State University, Institute for Research on Teaching.

Saigh, P. (1984). Unscheduled assessment: Test anxiety, academic achievement, and social validity. Educational Research Quarterly, 9(4), 6-11.

Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening? Phi Delta Kappan, 62, 631-634.

Sax, G., & Collett, L. S. (1968). An empirical comparison of the effects of recall and multiple-choice tests on student achievement. Journal of Educational Measurement, 5, 169-173.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38(3), 57-62.

Shavelson, R. J., Cadwell, J., & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. American Educational Research Journal, 14(2), 83-97.

Shulman, L. (1980). Test design: A view from practice. In E. Baker & E. Quellmalz (Eds.), Educational Testing and Evaluation. Beverly Hills, CA: Sage, 63-76.

Snyder, B. R. (1971). The Hidden Curriculum. Cambridge, MA: M.I.T. Press.

Sproull, L., & Zubrow, D. (1981). Standardized testing from the administrative perspective. Phi Delta Kappan, 62, 628-631.

Stetz, F. P., & Beck, M. D. (1981). Attitudes toward standardized tests: Students, teachers and measurement specialists. Measurement in Education, 12, 1-11.

Stiggins, R. J. (1985). Improving assessment where it means the most: In the classroom. Educational Leadership, 43, 69-74.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. Phi Delta Kappan, 69, 363-368.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practices, 5, 5-17.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. Journal of Educational Measurement, 26(3), 233-246.

Szafran, R. (1981). Question-pool study guides: Effects on test anxiety and learning retention. Teaching Sociology, 9, 31-43.

Takeuchi, R. T. (1977). Attitudes of elementary teachers toward testing: Use and abuse of standardized tests in California, 1976-77. Dissertation Abstracts International, 38, 4500A. (University Microfilms No. 7731120)

Trentham, L. (1975). The effect of distractions on sixth-grade students in a testing situation. Psychology in the Schools, 16, 439-443.

Tyler, R. W., & Sheldon, H. W. (1979). Testing, teaching, and learning. Report of a Conference on Research on Testing. Washington, DC: National Institute of Education.

Veenman, S. (1984). Perceived problems of beginning teachers. Review of Educational Research, 54(2), 143-178.

Wanous, D. S., & Mehrens, W. A. (1981). Helping teachers use information: The data box approach. Measurement in Education, 12(4), 1-10.

Wexley, K. N., & Thornton, C. L. (1972). Effect of verbal feedback of test results upon learning. Journal of Educational Research, 66, 119-121.

Yeh, J. P. (1981). Teachers and testing: A survey of test use. Washington, DC: U.S. Department of Health, Education and Welfare and National Institute of Education. (ERIC Document: ED 218 336)

9/7

Table 1

Teachers' Testing Practices, Attitudes, and Beliefs,

1. Teachers select and use assessment procedures that best fit their day to day instructional needs.

2. Teachers believe that in order for test results to be of use to them the tests must fit their instructional needs, must be of practical value, and must be immediately available.

3. Teacher-made tests are perceived by teachers to better meet their classroom instructional needs than are either standardized tests or state and school district pupil minimum competency tests.

4. Teachers rely on teacher-made tests to a much greater extent than standardized tests and district or state competency tests for making decisions about individual pupils.

5. Teachers believe that self-constructed assessments generally better meet the instructional needs of their classes than do assessments derived from other sources such as workbooks and textbooks.

6. Teachers believe that teacher-devised testing facilitates the classroom learning and teaching process.

7. Teachers believe, and indicate that school administrators and pupils also believe, that teacher-made tests should be scheduled on a relatively frequent basis to promote pupil learning.

8. Teachers believe that teacher-made test assessments should closely mirror instruction provided.

9. Teachers believe that teacher-made tests generally have a positive impact upon pupils and their study and learning efforts.

10. Teachers believe that teacher-designed testing and the discussion of test results following the testing sessions are productive uses of classroom time.

11. Teachers believe that differing course content and pupil grade level variations require somewhat different assessment devices and practices.

12. Teachers believe that the results from tests should be supplemented by information from other sources such as observations and daily work when assigning grades or making other decisions about pupils.

13. Teachers believe that daily classroom observations and teacher judgment are more reliable sources of information for making classroom related decisions than are isolated test scores.

(table continues)

Table 1 (continued)

14. Teachers believe that where student learning is displayed in overt behaviors less reliance should be made of paper and pencil type tests.

15. Teachers believe that test scores must be interpreted and used within the context of all other information available about a pupil.

16. Most teachers place considerable reliance on information about pupils gathered through informal observations, day to day communication, and daily work; teachers in the lower grades tend to rely more on these sources of information than on formal tests while middle and upper grade teachers tend to rely more on formal tests than upon informally gathered information.

17. Teachers believe that they are less proficient in testing skills when compared to their proficiencies in other professional skill areas.

18. Teachers believe that preservice training in tests and measurement provides them with adequate background concepts and principles but insufficiently prepares them for the successful integration of pupil assessment and instruction.

19. Teachers generally report that they have deficiencies in testing and measurement, feel that their self-constructed tests could be improved, and would like inservice training in tests and measurements if this training were oriented toward practical classroom needs, but they tend to be confident about their general testing abilities and knowledge.

20. Teachers believe that technical aspects of classroom testing such as use of test specification tables, item analysis procedures, test score statistical analyses, estimates of test reliability, and use of question writing guidelines are of limited practical value.

21. Teachers believe that teacher-made tests are useful in diagnosing pupils' progress, making pupil grouping decisions, assigning pupil grades, and reporting the progress of pupils.

22. Teachers believe that essay tests as compared to objective tests are impractical and disliked by pupils but result in greater study efforts and usually measure at higher cognitive levels.

23. Teachers believe that teacher-made test results aid teachers in justifying grades to pupils and parents.

24. Teachers believe that matching, short-response, completion, and multiple-choice questions are the more useable, efficient, and useful types of questions in contrast to the essay or true-false question types.

25. Teachers believe that testing and related assessment procedures, to be consistently used and useful in classrooms, must be efficient in time and energy demands of teachers and supportive of on-going classroom instructional activities.

(table continues)

BEST COPY AVAILABLE

Table 1 (continued)

26. Teachers believe that tests need to be administered fairly and efficiently and that testing periods should be monitored by teachers to prevent pupil cheating.

27. Teachers believe that test results can be interpreted and conveyed to pupils adequately without use of statistical analyses.

28. Teachers believe that a variety of question types should be used in classroom tests in order to be fair to pupils and better to complement various instructional objectives.

29. Teachers believe that teacher-made tests should contain questions that demand higher-order pupil thinking skills.

30. Teachers expend considerable class and work time and professional effort in testing and assessment activities, typically schedule formal tests once every two weeks or more often in most courses, construct on an average 54 formal tests each year, and construct most of their own test questions.

31. Teachers believe that testing, evaluation, and grading activities are among their more demanding and less pleasant classroom responsibilities.

32. Teachers commonly express concern about their pupil testing and evaluation responsibilities as well as about their class management and pupil motivation duties.

Table 2

Teachers' Testing Knowledge and Skills as Suggested by Perceptual Ratings of Their Testing Proficiencies, Tests of Their Knowledge, and Direct Analyses of Their Self-Constructed Tests

1. For most preservice and inservice teachers it appears that their knowledge of classroom testing practices and principles is inadequate to meet classroom evaluation needs, and it appears that little progress has been made in overcoming this inadequacy during the past quarter century.

2. In more recent studies teachers' performance on paper and pencil measures of knowledge of classroom testing concepts and principles still appears to be in the 50 percent correct range as was found in Mayo's classic study in 1967. Some researchers have estimated that no more than 25 percent of K-12 classroom teachers can correctly answer basic questions on classroom measurement concepts and principles.

3. Many practicing teachers report having received no formal measurement training during their preservice education, many practicing teachers report having received just a single unit of instruction in measurement within another preservice education course, and most practicing teachers report having received no school sponsored inservice training in the development and use of classroom tests.

4. Neither inservice training, if provided, nor increased years of teaching experience appear to improve either classroom teachers' testing knowledge or their test construction skills as measured by paper and pencil tests and as revealed by direct analyses of construction flaws found on their self-constructed tests.

5. School principals and supervisors rate beginning teachers' testing proficiencies lower than they rate beginning teachers' proficiencies in other professional areas; practicing teachers also rate their testing proficiencies lower than they rate their professional proficiencies in other skill areas.

6. Teachers' with typical formal training in tests and measurement perform better than teachers without this training on paper and pencil measures of testing knowledge, but their scores typically exceed the scores of untrained teachers by just six to 10 percent.

7. Teachers have difficulty in correctly answering questions related to appropriate interpretations of scores commonly used in conveying pupil performance on standardized and state competency tests.

8. Principals and supervisors perceive beginning teachers and experienced teachers perceive themselves to have lower proficiencies in conducting simple statistical analyses of test scores, in writing questions demanding higher thinking skills, and in use of sociometric techniques then compared to their proficiencies in test planning, interpretation, and use.

(table continues)

Table 2 (continued)

9. Teachers display especially limited knowledge about technical aspects of testing (e.g., use of test specification tables, item analysis and statistical analysis procedures, etc.).

10. Analyses of teachers' tests reveal very frequent violations of common question and format construction guidelines with matching exercises being found to be particularly error prone.

11. Teachers tend to frequently use short-answer, completion, and matching question types which commonly measure at the lower cognitive demand levels. Multiple-choice questions are also commonly used; true-false are used less often; and essay questions are used very infrequently.

12. Teacher-constructed tests measure predominantly at the knowledge cognitive functioning level (approximately 70 to 100 percent of all items on their tests) with more higher level functioning items typically found on math and science tests and with all test items used in social studies and other subject areas functioning almost exclusively at the knowledge level.

13. Many teacher-constructed tests are reported to be almost illegible due to poor typing or poor handwriting, lack of concern about format, and/or poor duplication quality.

14. The types of test questions used by teachers vary somewhat by subject area, content being assessed, and grade level of instruction.

15. Teacher-constructed tests typically contain approximately 35 questions with an average of 2.6 different question types being used.

16. Many teacher-made tests contain incomplete, inadequate, or completely lack directions.

17. Teachers appear to be unable to identify common test question construction guideline flaws or violations on their tests and report spending little time editing or revising test questions. Some indirect evidence suggests that school principals and supervisors also are unable to distinguish between poorly and well written test question exercises.

18. Teachers appear to value the importance of having higher cognitive functioning questions on teacher-made tests, but they infrequently use such questions; they tend to over-estimate the number of higher order questions used on their tests; and they have difficulty identifying and writing test questions that function beyond the knowledge level.

19. Teachers, principals, and supervisors rate teachers' grading related skill proficiencies higher than they rate teachers' test item writing proficiencies.

(table continues)

BEST COPY AVAILABLE

Table 2 (continued)

20. Teachers, principals, and supervisors appear to agree rather highly one with another about the relative level of teachers' proficiencies in various testing skills; they also agree one with another that teachers' preservice preparation in testing is less adequate than their level of preparation in other areas of professional training.

21. Teachers', principals', and supervisors' ratings of the levels of teachers' proficiencies in writing various test question types are highly, but negatively, correlated with the levels of the frequencies of violations of item construction guidelines found in teacher-made tests.